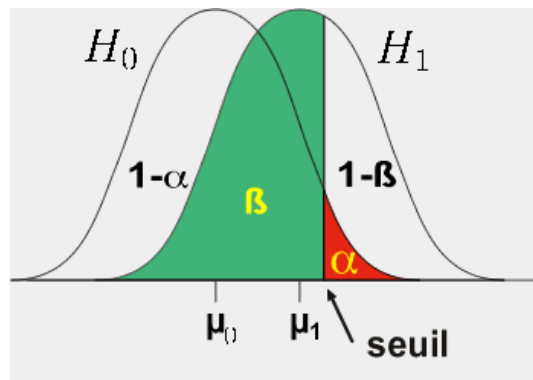


---

# Statistique - Tests d'hypothèses

## Exercices

---



Fabrice Heitz

Septembre 2013

---

# 1 Tests sur une seule population (comparaison par rapport à une référence)

---

## Exercice 1 : Test sur la moyenne : réglage d'une machine

1. Une machine produit des tiges métalliques dont la longueur nominale est égale à 8,30 cm. Les fluctuations de longueur dues au procédé de fabrication correspondent à un écart-type de 0,6 cm. Sur la base d'un échantillon aléatoire de taille  $N = 100$ , on veut tester si la machine est bien réglée. La moyenne des longueurs mesurées sur l'échantillon est de 8,57 cm.

Faut-il réaliser un test unilatéral ou bilatéral ? Conclure pour un seuil de signification  $\alpha$  de 5%, 1% et 0,1%.

*Réponse :  $Z = 4,50$ . L'hypothèse  $H_0$  (machine bien réglée) est rejetée pour toutes les valeurs de  $\alpha$ .*

2. Quelle serait la conclusion avec seulement 20 échantillons pour la même moyenne mesurée (on supposera dans ce cas que la longueur des pièces produites suit une loi normale) ?

*Réponse :  $Z = 2,01$ . L'hypothèse  $H_0$  est rejetée pour  $\alpha=5\%$ . On ne peut rejeter  $H_0$  pour  $\alpha=1\%$  et  $0,1\%$ .*

3. Une machine produit des tiges métalliques dont la longueur nominale, égale à 8,30 cm est supposée suivre une loi normale. Sur la base d'un échantillon aléatoire de taille  $N = 20$ , on veut tester si la machine est bien réglée.

La moyenne et l'écart-type des longueurs **mesurées sur l'échantillon** est respectivement de 8,57 m et 0,6 cm. Que peut-on en conclure (pour les mêmes seuils de signification) ?

*Réponse :  $T = 2,01$ . L'hypothèse  $H_0$  n'est rejetée pour aucune valeur de  $\alpha$ .*

4. Comparer et commenter les différents cas.

## Exercice 2 : Test sur la variance : précision d'usinage de pièces automobiles

Les pièces des moteurs d'automobiles de dernière génération sont usinées avec une très grande précision. L'écart-type des dimensions d'une pièce ne doit pas dépasser  $10\mu m$  (les dimensions suivent une loi normale, la dimension moyenne n'est pas donnée).

On prélève sur une unité de production 25 moteurs pour lesquels des mesures dimensionnelles menées sur la même pièce donnent un écart-type de  $13,5\mu m$ . Peut-on conclure que la

variabilité est supérieure à la valeur tolérée (on adoptera un seuil de signification de 1%) ?  
*Réponse :  $\chi^2 = 43,7$ . Rejet de  $H_0$  au seuil 1%. Donc on peut conclure que l'unité de production est défectueuse.*

### Exercice 3 : Test sur une proportion : la marche des bébés

Des études en psychologie du développement ont montré qu'à l'âge de 12 mois, 50 % des bébés « normaux » marchent.

On souhaite mener une étude sur les retards de développement des bébés prématurés. On teste l'hypothèse que les bébés prématurés marchent plus tardivement que les bébés normaux.

On observe une population de 80 bébés prématurés. A 12 mois, 35 de ces 80 bébés marchent.

Faut-il réaliser un test unilatéral ou bilatéral ? Peut-on, au seuil de signification de 5%, valider l'hypothèse de recherche ? Quelle est la p-valeur dans ce cas ?

*Réponse :  $Z = -1,12$ . On ne peut rejeter  $H_0$  (les observations disponibles ne montrent pas que les bébés prématurés marchent plus tardivement). La p-valeur est :  $p = 0,1314$  (donc le test serait significatif pour un choix de  $\alpha > 13,14\%$ ).*

### Exercice 4 : Test sur une proportion : le dé est-il pipé ?

On joue avec un dé qui semble tomber *trop souvent* sur la face 6. Dans une expérience on a lancé 40 fois ce dé et obtenu 10 fois le 6.

1. Au seuil de signification de 5% peut-on conclure que le dé est pipé par excès d'apparition du 6 ?
2. Sur 40 lancers, à partir de quelle proportion de faces 6 peut-on conclure que le dé est pipé (au seuil de signification de 5%) ?
3. Quelle est la réponse à la question 2, lorsqu'on ne sait pas a priori si le dé, que l'on soupçonne d'être pipé, tombe trop souvent ou pas assez souvent sur la face 6 ?

*Réponses : 1)  $Z = 1,41$ . On ne peut donc conclure que le dé est pipé. 2) Pour conclure à un dé pipé on devrait observer, sur 40 lancers, un nombre de « faces 6 »  $\geq 11$  (test unilatéral). 3) Dans ce cas le nombre de « faces 6 » doit être  $\geq 12$  ou  $\leq 2$ .*

---

## 2 Tests comparant deux populations

---

### Exercice 5 : Comparaison de moyennes : pH-mètre

1. Pour mesurer le pH d'une solution, on utilise un pH-mètre qui affiche un résultat dont la loi est  $\mathcal{N}(\mu, 0,05^2)$ , où  $\mu$  est la vraie valeur du pH de la solution. On a mesuré le pH d'une solution A par 12 mesures indépendantes et trouvé une moyenne de 7,4, et le pH d'une solution B par 10 mesures indépendantes et trouvé une moyenne de 7,5. Peut-on considérer que les deux solutions ont même pH, au niveau de risque 1% ?

*Réponse :  $Z = 4,67$ . L'hypothèse  $H_0$  est rejetée au seuil 1% (et même 0,1%). On conclut donc que les deux solutions n'ont pas même PH.*

2. Pour mesurer le pH d'une solution, on utilise un nouveau pH-mètre qui affiche un résultat dont la loi est  $\mathcal{N}(\mu, \sigma^2)$ , où  $\mu$  est la vraie valeur du pH de la solution et où  $\sigma$  n'a pas été déterminé. On a mesuré le pH d'une solution A par 12 mesures indépendantes et trouvé une moyenne de 7,4 et un écart-type empirique de 0,09, et le pH d'une solution B par 10 mesures indépendantes et trouvé une moyenne de 7,5 et un écart-type empirique de 0,08. Peut-on considérer que les deux solutions ont même pH ?

*Réponse :  $T = 2,73$ . les mesures disponibles ne permettent pas de rejeter  $H_0$  (solutions de mêmes PH), au seuil 1%*

### Exercice 6 : Comparaison de moyennes et variances : taille des poissons

*d'après Stéphane Guidon, DESS bioinformatique, LIRMM*

Dans une étude en biologie on mesure la longueur des spécimens mâles et femelles de poissons adultes appartenant à la même espèce. On obtient les résultats suivants (longueurs en mm) :

mâles      120   107   110   116   114   111   113   117   114   112

femelles   110   111   107   108   110   105   107   106   111   111

Peut-on affirmer que la taille des individus diffère entre les deux sexes de cette espèce ? (On supposera que les tailles dans les deux populations sont distribuées selon des lois normales). On n'oubliera pas de tester tout d'abord l'hypothèse d'homoscédasticité (égalité des variances). On prendra  $\alpha = 5\%$ .

Réponse :  $N_1 = 10$  ;  $\bar{X}_1 = 113,40$  ;  $S_1 \simeq 3,7178$  ;  $N_2 = 10$  ;  $\bar{X}_2 = 108,60$  ;  $S_2 \simeq 2,2706$   
 Test d'égalité des variances :  $F \simeq 2,68$ , seuil de décision pour  $\alpha = 5\%$  :  $4,03$ . Donc l'hypothèse d'égalité des variances est acceptée. On peut donc réaliser le test de comparaison des moyennes (test de Student) :  $T \simeq 3,48$ , seuil de décision pour  $\alpha = 5\%$  :  $2,10$ . Conclusion : les tailles sont en moyenne significativement différentes.

## Exercice 7 : Comparaison de proportions : traitements anti-cancéreux

On teste deux traitements anti-cancéreux  $A$  et  $B$  sur deux populations de patients  $\mathcal{P}_A$  et  $\mathcal{P}_B$  (de même taille  $n_A = n_B = 50$ ). L'efficacité d'un traitement est évaluée par l'éventuelle diminution de la taille de la lésion tumorale, estimée par imagerie médicale, après un an de traitement.

Pour la population soumise au traitement  $A$  on observe une diminution de la taille des tumeurs dans 27 cas sur 50, pour le traitement  $B$ , dans 18 cas. Peut-on conclure à une différence d'effet des deux traitements (au seuil de 5 %) ? Peut-on conclure que le traitement  $A$  est plus efficace que le traitement  $B$  (avec le même seuil de signification) ?

Réponse : valeur du test =  $1,8091$

Test bilatéral : seuil  $1,96$ , on ne peut rejeter  $H_0$

Test unilatéral : seuil  $1,64$ , rejet de  $H_0$

## 3 Tests du $\chi^2$

### Exercice 8 : Test d'homogénéité : comparaison de méthodes pédagogiques

d'après F. et M. Diener, Université Nice Sophia Antipolis

Dans une Ecole, trois groupes de professeurs ont mis au point trois méthodes différentes d'enseignement des statistiques, qu'on a appliqué à trois échantillons d'étudiants ayant sensiblement le même niveau initial. A l'examen les résultats furent les suivants :

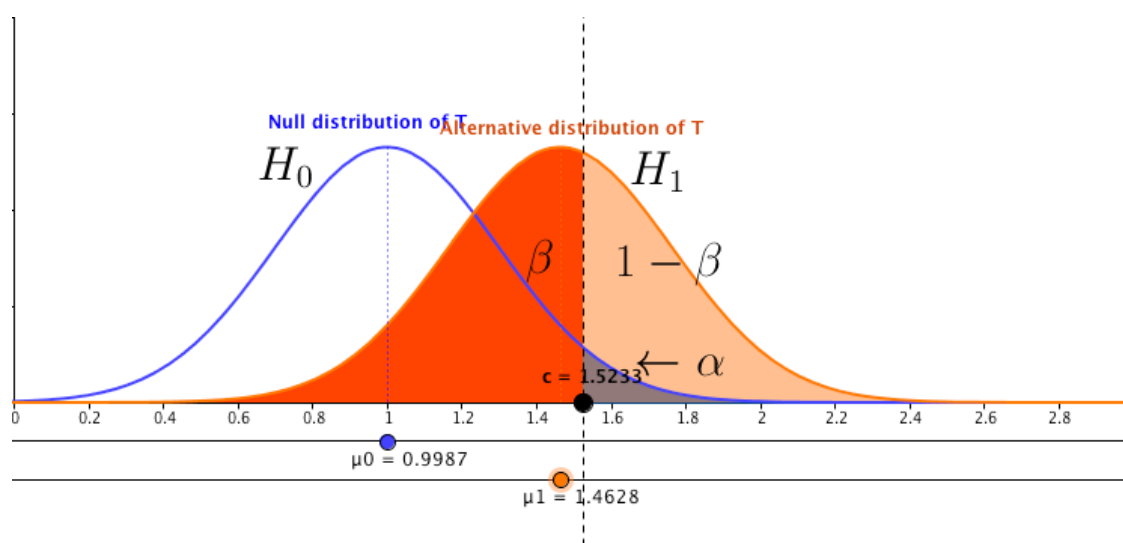
	Admis	Ajournés	Total
Méthode 1	51	29	
Méthode 2	38	12	
Méthode 3	86	34	
Total			

Peut-on affirmer que l'une des trois méthodes est plus efficace que les autres en termes de réussite à l'examen ? (Effectuer un test d'homogénéité).

Réponse :  $\chi^2 \simeq 2,51$ ,  $ddl=2$ , à  $\alpha = 5\%$  on ne peut rejeter  $H_0$ .

## 4 Calcul de la puissance d'un test

### Exercice 9 : Calcul de puissance dans un test (comparaison d'une moyenne à une référence)



Applet : <http://www.math.usu.edu/~schneit/CTIS/HTErrors/>

FIGURE 1 – Les deux types d'erreurs dans un test d'hypothèse

	$H_0$ vraie	$H_1$ vraie
rejet de $H_0$ (décision $H_1$ )	$\alpha$	$1 - \beta$
non rejet de $H_0$	$1 - \alpha$	$\beta$

- **Erreur de 1ère espèce** : décider  $H_1$  (rejeter  $H_0$ ) alors que  $H_0$  est vraie  
 $\alpha = p(H_1|H_0)$  risque de 1ère espèce
- **Erreur de 2ème espèce** : ne pas rejeter  $H_0$  alors que  $H_1$  est vraie  
 $\beta = p(H_0|H_1)$  risque de 2ème espèce
- **Puissance du test** :  $1 - \beta = p(H_1|H_1)$

La puissance mesure la capacité du test à discriminer correctement les deux hypothèses. Un bon test est un test qui, pour  $\alpha$  donné<sup>1</sup>, maximise la puissance  $1 - \beta$  (ou minimise le risque de 2ème espèce  $\beta$ ). Un test satisfaisant doit avoir une puissance d'au moins 0,80.

1. on choisit souvent  $\alpha = 5\%$  ou  $\alpha = 1\%$

**Enoncé de l'exercice :** on considère un test de comparaison de la moyenne d'une variable d'intérêt sur une population (e.g. durée de vie d'un moteur, voir étude de cas du cours) par rapport à une valeur de référence  $\mu_0$ . La variable aléatoire d'intérêt  $X$  est supposée suivre une loi normale  $\mathcal{N}(\mu, \sigma)$ . Les deux hypothèses à tester sont :

- $H_0 : \mu = \mu_0$
- $H_1 : \mu > \mu_0$  (test unilatéral)

On note  $\mu_1$  la *vraie* moyenne si  $H_1$  est vraie.

Le test est mené sur  $N$  échantillons  $X_i$  prélevés sur la population.

Le test est mené sur la moyenne empirique :  $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$

- si  $H_0$  est vraie :  $\bar{X} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{N})$
- si  $H_1$  est vraie :  $\bar{X} \sim \mathcal{N}(\mu_1, \frac{\sigma^2}{N})$

La statistique  $\mathcal{T}$  utilisée pour le test est :  $\mathcal{T} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}}$

On se fixe le risque de 1ère espèce  $\alpha$  et on détermine le seuil de décision  $z_\alpha$  tel que :  
 $p(z > z_\alpha) = \int_{z_\alpha}^{+\infty} f_Z(z) dz = \alpha$  où  $z \sim \mathcal{N}(0, 1)$  (pour  $\alpha = 5\%$ ,  $z_\alpha = 1,64$ ).

1. Montrer que l'expression de la puissance  $1 - \beta$  pour ce test est :

$$1 - \beta = p\left(z > \underbrace{z_\alpha - \left(\frac{\mu_1 - \mu_0}{\sigma}\right)\sqrt{N}}_{z_{1-\beta}}\right)$$

où  $z \sim \mathcal{N}(0, 1)$ .

On note  $\Delta = \mu_1 - \mu_0$  la différence entre les moyennes sous  $H_1$  et  $H_0$ .

2. Etudier qualitativement les variations de la puissance du test lorsque :
  - $\alpha$  varie
  - $\Delta$  varie
  - $N$  varie

*Réponse : lorsque  $\alpha \searrow$ ,  $\beta \nearrow$  et la puissance  $\searrow$*

*lorsque  $\Delta \nearrow$ ,  $\beta \searrow$  et la puissance  $\nearrow$*

*lorsque  $N \nearrow$ ,  $\beta \searrow$  et la puissance  $\nearrow$*

On utilisera les résultats des calculs précédents et la simulation disponible sur :

[http://wise.cgu.edu/power\\_applet/power.asp](http://wise.cgu.edu/power_applet/power.asp)

3. Pour  $\alpha = 0,05$ ,  $\sigma = 1$  et  $N = 30$ , déterminer la plus petite différence de moyenne  $\Delta$  que l'on peut mettre en évidence avec ce test, en garantissant une puissance  $1 - \beta \geq 0,80$ .  
Même question pour  $N = 10$  et  $N = 20$ .

*Réponse :  $N = 10 : \Delta_{min} \simeq 0,78$ ,  $N = 20 : \Delta_{min} \simeq 0,55$ ,  $N = 30 : \Delta_{min} \simeq 0,45$*

4. Quel est le nombre minimal d'échantillons nécessaires pour mettre en évidence une différence de moyenne de  $\Delta = 0,1$ ? (en conservant une puissance  $\geq 0,80$ ).

*Réponse :  $N \geq 616$*

On utilisera les résultats des calculs précédents et la simulation disponible sur :

[http://wise.cgu.edu/power\\_applet/power.asp](http://wise.cgu.edu/power_applet/power.asp)